

# A new surveillance and spatio-temporal visualization tool SIMID: SIMulation of Infectious Diseases using random networks and GIS

Lilia L. Ramírez-Ramírez<sup>a,\*</sup>, Yulia R. Gel<sup>b,c</sup>, Mary Thompson<sup>b</sup>,  
Eileen de Villa<sup>d</sup>, Matt McPherson<sup>e</sup>

<sup>a</sup> Instituto Tecnológico Autónomo de México (ITAM), Mexico City, Mexico

<sup>b</sup> Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

<sup>c</sup> Department of Applied Mathematics and Statistics, Johns Hopkins University, USA

<sup>d</sup> Peel Public Health, 7120 Hurontario Street, Mississauga, ON, Canada

<sup>e</sup> Infonaut Inc., 255 Consumers Road, Suite 500, Toronto, ON, Canada

## ARTICLE INFO

### Article history:

Received 7 February 2012

Received in revised form

24 October 2012

Accepted 11 January 2013

### Keywords:

Software for epidemic surveillance

Disease outbreaks

Spatio-temporal models in  
epidemiology

Networks

## ABSTRACT

In this paper we discuss the SIMID tool for simulation of the spread of infectious disease, enabling spatio-temporal visualization of the dynamics of influenza outbreaks. SIMID is based on modern random network methodology and implemented within the R and GIS frameworks. The key advantage of SIMID is that it allows not only for the construction of a possible scenario for the spread of an infectious disease but also for the assessment of mitigation strategies, variation and uncertainty in disease parameters and randomness in the progression of an outbreak. We illustrate SIMID by application to an influenza epidemic simulation in a population constructed to resemble the Region of Peel, Ontario, Canada.

© 2013 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

The prompt detection of respiratory and gastrointestinal infectious diseases is of critical interest in public health practice due to the rapid transmission of such illnesses and their potential burden on the community. There is growing evidence that dynamic space–time data constitute a key element in monitoring and forecasting the spread of certain infectious diseases (see, for example, reviews by [25,38,6,42,19]). In particular, the threat of a severe influenza pandemic raises

significant concern which necessitates the early spatial detection of an outbreak for appropriate public health action [30,9,20,27,44]. However, traditionally most public health departments do not incorporate space–time dynamic information into disease surveillance, reducing the efficiency and speed with which outbreaks can be detected and interventions started. Hence, there exists a substantial demand for easy-to-use, map-based software and decision-making support services that could assist healthcare professionals in assessing and integrating actionable space–time information on infectious diseases.

\* Corresponding author. Tel.: +52 55 5628 4000.

E-mail address: [lilialeticia.ramirez@itam.mx](mailto:lilialeticia.ramirez@itam.mx) (L.L. Ramírez-Ramírez).  
0169-2607/\$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.  
<http://dx.doi.org/10.1016/j.cmpb.2013.01.007>

The classical tool for visualizing spatial information on disease spread is a geographic information system (GIS). However, tracking the dynamics of infectious diseases and detecting changes in disease processes are impossible without the development and implementation of statistical methodology for spatio-temporal disease surveillance, which is not routinely included in GIS and most statistical software packages. There exists a vast literature on statistical approaches for space-time epidemiology [45,25,48,7,43], which can be generally classified into the three main groups, namely, statistical tests, model-based methods and hybrid routines. The first group, which includes, for example, such methods as scan statistics and cumulative sums, is dominantly employed for detecting outbreaks by comparing a spatio-temporal subset of data vs. the expected rate of disease incidence [13,23,24]. The second group focuses on assessing the relative risk of disease occurrence within a domain of interest, with respect to various spatio-temporal environmental and demographic covariates. Some of the most widely used approaches in this group are spatio-temporal Bayesian methods [47,26] and references therein and generalized linear models [22]. The hybrid group of surveillance methods includes modern simulation-based routines that aim simultaneously to model and to detect the dynamics of infectious diseases, and includes such approaches as hidden Markov models [49], Bayesian nets [18], agent-based models [11,10] and references therein and network methodology [35,36,4,2,21,17], and references therein. As mentioned by [42], while many of these new hybrid methods appear to be promising, they are not yet broadly implemented in software and, hence, are not widely available for operational disease monitoring.

In this paper we describe a new simulation-based visualization program SIMID for infectious disease surveillance, based on modern network methodology<sup>1</sup> and implemented within the R and GIS frameworks [8]. Our new network epidemic model adapts and extends the models for networks of social contacts proposed by [35,4], by allowing us to consider not only a single random network that is described by a specific degree distribution, but also a population network of contacts made up of subnetworks. Moreover, each such subnetwork can follow a different degree distribution. The network describes the contacts that can result in disease transmission, and is defined in terms of a probability distribution for the number of contacts that each susceptible individual has with other individuals in the community. Moreover, the network methodology we use for disease dynamics generalizes some important epidemic models, relaxing the hypotheses that the population is homogeneously mixed, that the infection rate is constant between any two individuals, and that the latent and infectious period have lengths

that are exponentially distributed [39–41]. Hence, the network algorithm incorporates not only the use of compartments as in [32,31,12] but also some of the most important everyday contacts that can be modeled through random networks, such as contacts within families, educational facilities, health care institutions and transportation hubs. In addition, our new surveillance simulation-based tool allows calculation of probability distributions for the total number of infected individuals and the replacement number<sup>2</sup> under four different control measures: mass vaccination and acquaintance vaccination that are implemented prior to the outbreak, and ring vaccination and isolation that are applied during the outbreak evolution. Thus, the SIMID tool enables not only the modelling of a single potential scenario of an epidemic outbreak but rather the generation of an ensemble of potential scenarios for infectious outbreaks and, hence, the assessment of various associated uncertainties and mitigation strategies. The developed methodology and software tool are applicable to any population with a possibly heterogeneous contact structure described as a random graph. The spatial view of social interactions that represent channels of diseases transmission and that are modeled with the random network methodology is placed within the GIS framework. This, along with the uncertainty quantification techniques, enables surveillance teams and other health care professionals to identify disease outbreaks in a prompt time frame (if the SIMID tool is initialized with currently observed data on influenza-like illness (ILI) occurrences or confirmed influenza infected individuals), and allows public health personnel to use the program for planning preventive interventions and training (if the SIMID tool is run based on simulated data). Although our current case study relates to visualization over time and space of the dynamics of influenza, including pH1N1 and seasonal influenza, the simulation tool is also applicable to a wide range of other infectious diseases, such as measles, meningococcal meningitis, enteric illnesses and sexually transmitted diseases, to name only a few.

The paper is organized as follows. In Section 2, we describe the process of generating the network of contacts, in its application to an influenza epidemic simulation in the Region of Peel, Ontario, Canada, taking into account demographic information from the 2006 census. In Section 3, we discuss the SIMID tool along with its most important methodological characteristics. Section 4 is devoted to technical software specifications and architecture for SIMID. Section 5 provides information on the workflow for building possible scenarios for infectious outbreaks and presents a sample run of the SIMID tool. The paper is concluded by discussion and future work.

<sup>1</sup> Throughout the text, “network model” refers to a specific population’s network of contacts, while “network algorithm” refers to selection of subpopulations (using available information, e.g. census data) and the computational algorithm that is used for generation of subnetworks. In turn, “network epidemic model” refers both to the network model and the epidemic model spreading along this network.

<sup>2</sup> Here by replacement number we mean the expected number of new cases generated by one single infected secondary case [5,35]. Note that the basic reproduction number and the replacement number are generally not the same since the basic reproduction number describes the expected number of new cases generated by patient zero, i.e. the first infected individual. It can easily be shown that when the network of contacts is modeled as a random network with Poisson degree distribution, then the reproductive and replacement number coincide.

## 2. Methods: network of contacts

The key idea of our simulation tool is to describe a population by a network whose vertices are individuals or groups of individuals and whose edges represent the kind of contacts that an infectious agent can use for transmission from an infective to a susceptible individual. Hence, as an input, our outbreak simulation program utilizes network population contact information that is grouped into two categories, namely, the network's vertices and the edges connecting them. The two input files are named respectively **indexes.txt** and **edges.txt**, and in this section we describe the general procedure to generate both files, using census data and information on the location of schools and their catchment areas.

For the specific case of the model population resembling that of the Region of Peel, the vertices in the contact graph are all individuals who are part of at least one of three kinds of subnetworks: “households”, “schools” and “general adult”. The first two classes of subnetworks have as many elements as there are households and schools in the community, respectively.

The elements in the file **indexes.txt** are the individuals in the population. They are characterized with an ID number. This file also stores individual information such as sex, age group (or age) and the subnetworks in which they have connections.

In the file **edges.txt** we indicate the end vertices of each edge without considering any direction. Then the edge that connects vertices  $v_1$  and  $v_2$  is included as  $v_1, v_2$  if  $v_1 < v_2$  or as  $v_2, v_1$  if  $v_2 < v_1$ . This file also indicates in which subnetwork each edge occurs. For example, if individuals  $v_1$  and  $v_2$  are from different dwellings but they are connected at school  $j$ , then this edge is characterized as a  $j$ -school subnetwork edge.

In order to construct the model population's network of contacts, we used information from the 2006 Canadian census [46] tabulated by Forward Sortation Areas (FSAs) and municipalities. Based on this information and also on the school locations and their catchment areas, we constructed the household, school and adult general connectivity of the population. Since the level of information that these sources provide does not describe the population connectivity completely, we introduced random elements within the subnetworks. Thus we completed the network to the required level of detail, but incorporating the uncertainty in our knowledge of the population. For example, in order to reflect the chance nature of social interaction, the contact in the general adult subnetwork has edges that are randomly generated according to the user-selected probability degree distribution.

In the census information we find the counts of individuals (sex, age), households (size and structure – number of members and age composition), and dwellings (number of households). With these data, dwellings in the Region of Peel were assigned to locations, and households were assigned to dwellings, and families were allocated (by structure and parents' status) to households; then individuals were allocated to families.

Based on the information from schools, the families were then assigned to school catchment areas, and with this information their children were placed into schools.

**Table 1 – Number of occupied private dwellings by municipality (2006 census).**

Brampton	125,934
Caledon	18,214
Mississauga	214,894
Total Peel Region	359,042

To describe the interactions of in-school children and of outside-dwelling adults two classes of subnetworks were considered. These subnetworks were modeled as random networks.

In the following subsections we describe with more detail the process of constructing the household-school networks and the in-school and outside-dwelling networks.

### 2.1. Households

#### 2.1.1. Dwellings

Using the variable “total number of occupied private dwellings” in the 20% sample data from the 2006 Census (Table 1), we obtained the total number of dwellings within each dissemination area. (Each of these areas is composed of one or more neighboring dissemination blocks, with a population of 400–700 persons (Statistics Canada).) Since the information is rounded to 5 units, the total number of dwellings was normalized to agree with the 2006 census total number of occupied private dwellings by municipality (Table 1).

The dwellings within each dissemination area were then assigned to a longitude–latitude location within the area delimited by the dissemination area borders, according to a uniform random distribution within a polygon area. (This information, together with the school catchment areas provided to us as polygons, was subsequently used to define which school each child attends.) In particular, to assign dwellings we employ the ArcGIS extension method XTools that creates uniform random sampling of points within a polygon describing the dissemination area. XTools is one of the most popular independent (from ESRI) scripts and is currently used by many governmental institutions including the National Oceanic and Atmospheric Administration (NOAA).

Based on census information, Statistics Canada also reports the number of households by familial structure:

- 1 a one-family household.
- 2 a multiple-family household.
- 3 a non-family household.

Since this familial information is provided for each dissemination area, the dwellings in each area are distributed among the three classifications of household. Due to the fact that there is no further information about the spatial distribution of any of these types of households within dissemination areas, the already spatially allocated dwellings in each area are randomly classified by simple random sampling, successively for each household type.

According to Statistics Canada, a family is defined as consisting of individuals that live in the same dwelling and are: a couple (married or common law) and the children, if any, of

either or both spouses; or, a lone parent of any marital status with at least one child living in the same dwelling and that child or those children. A couple may be of opposite or the same sex. Children may be children by birth, marriage or adoption regardless of their age or marital status as long as they live in the dwelling and do not have their own spouse or child living in the dwelling. Grandchildren living with their grandparent(s) but with no parents present also constitute a census family.

### 2.1.2. Families

Using the information from the variable “Total number of census families in private households” by the status of “Married couple”, “Common law couples”, “Female parent” and “Male parent” in the 20% sample data from the Canadian census, the families (by dissemination area) were assigned to one of the four “couple/parents’ status” categories:

- 1 Married couple (with or without children).
- 2 Common law couple (with or without children).
- 3 Single female parent family.
- 4 Single male parent family.

Multiple-family households were considered to have two families, since in the census data only 5% of dwellings have three or more families. In particular, we allocate families per dwellings with probabilities that are proportional to their numbers in the available census data. For instance, in the current Peel Region data, the proportion of dwellings with one family is approximately 89%, proportion of dwellings with two families is 6% and proportion of dwellings with three or more families is 5%. Hence, given that dwellings with two families and dwellings with three or more families are rare, i.e. only 6% and 5% of all dwellings respectively, and since we currently have no more specific information on proportion of other multiple-family dwellings, we combine dwellings with two and more families into one category. (Similarly, if, for example, in some other data set we observe that 6% of dwellings have two or more families, then we can either assume that all these 6% of dwellings have two families, if no extra information is available; or if additional information on multiple-family households is available, we incorporate it proportionally to a number of dwellings with three, four etc families.)

From the census municipality information on number of children by parental status, we randomly assigned a number of children to each family, according to the following procedure of successive simple random sampling. Suppose that there are  $N$  families and  $N_i$  are assumed to have  $i$  children,  $i = 0, 1, 2, \dots, m$ , then we select a simple random sample without replacement of  $N_0$  families to have 0 children; then among those that are left, a simple random sample without replacement of  $N_1$  to have 1 child; and so on. Since the information is presented aggregated as: “without children”, “one child”, “two children” and “three or more children”, in this step, numbers of children were assigned to families assuming that families with “three or more children” had exactly three children.

**Remark.** Given the currently available census data, this assumption is made for simplicity, but it might be too restrictive since large families play a significant role in influenza epidemics. Hence, there are three options to relax this

assumption. First, a user can employ a different empirical data source for the family size distribution. (Unfortunately, we currently have no access to other data sources for Peel Region beyond the Canadian census data.) Second, the code for constructing the overall population network offers an option to approximate family size distribution with a Poisson distribution, i.e. using information on proportions of families with 1 and 2 children, we can infer proportions of families with 3, 4 etc. number of children. Finally, a user is offered an option to allocate children from families with more than three children, following a two-stage allocation algorithm. With this method, in the first stage we allocate all children who are to belong to families with one and two children to families with those compositions, and in the second stage we (uniformly randomly) assign the remaining children (4% in the case of Peel Region) into families with “at least three children”.

Then, for each family we have the following information:

- 1 type.das: Type of dwelling (one-family, multiple-family or non-family).
- 2 dwell.id: Occupied dwelling identification number.
- 3 fam.id: Family identification number within the dwelling. It has the value 1 if the family is in a one-family household or if it is the first family in a multiple-family household. It has the value 2 if it is the second family in a multiple-family household.
- 4 fam.str: Parents’ status, with value 1, 2, 3 or 4.
- 5 n.child: The number of children in the family.

### 2.1.3. Family members

Based on the 2006 census tables of “Household Living Arrangements” (HLA), available for the Region of Peel at the website of Statistics Canada (Household Living Arrangements, Age Groups and Sex for the Population in Private Households of Canada, Provinces, Territories, Census Divisions and Census Subdivisions, 2006 Census – 20% Sample Data), we assigned each couple or single parent to one age-sex group, according to the following procedure.

The tables used were specific to municipality and sex, and they presented the total number of inhabitants by age group and subtotals according to the following categorizations:

- 1 Total.
- 2 Total persons in family households.
- 3 Spouses, common law partners or lone parents.
- 4 Sons and/or daughters in families with two parents present.
- 5 Sons and/or daughters in families with one parent present.
- 6 Persons not in families, living with relatives (non-relatives may be present).
- 7 Persons not in families, living with non-relatives only (these non-relatives must constitute a census family).
- 8 Total persons in non-family households.
- 9 Living with relatives (non-relatives may be present).
- 10 Living with one or more non-relatives only.
- 11 Living alone.

The age group selection for an adult was weighted according to the observed age group distribution crossed with the variable “Spouses, common law partners or lone parents”



**Table 2 – Age groups. Note that age group 14 is a ten-year rather than a five-year age group.**

Group	Years
1	0–4
2	5–9
3	10–14
4	15–19
5	20–24
6	25–29
7	30–34
8	35–39
9	40–44
10	45–49
11	50–54
12	55–59
13	60–64
14	65–74
15	75–79
16	80–84
17	85–100

(male and female tables). The procedure was then implemented iteratively considering the remaining population to allocate, that is, the weight for a group was decreased after assigning membership in the group.

If the parents' status was married or common law, then the male's age was chosen by weighted random selection from any available age, and the female parent's age group was chosen similarly, but only from those age groups that were between two groups below and one above the male's.

The children's age groups (Table 2) were obtained from the age distributions for "Sons and/or daughters in families with two parents present" and "Sons and/or daughters in families with one parent present" from the same HLA tables. As in the parents' case, the allocation process was iterative and restricted by the parents' and already allocated siblings' age groups.

The first child's age was randomly selected from the age groups at most equal to the age group three groups below the oldest parent's age, first according to a uniform distribution among age groups and then according to a uniform distribution among ages within the selected age group. If the family had two children, then the age group for the second child was drawn from the groups with the same restriction as before but also from the age groups that are 4 age groups around (above or below) his/her sibling. Thus, if a family had three children and the age group of the oldest parent was  $a$ , the first child age group ( $b_1$ ) could be between 1 and  $a - 3$ . The second child could have age group ( $b_2$ ) between  $\max(1, b_1 - 4)$  and  $\min(a - 3, b_1 + 4)$ , and the third child could have age group  $\max(1, \min(b_1, b_2) - 4)$  to  $\min(a - 3, \max(b_1, b_2) + 4)$ .

Since the HLA files also provide information on the persons living in families by age group, municipality and sex, then the previous procedure can be applied to allocate the children in family not only by age group, but also by sex.

Since we considered that families with at least three children had exactly three children, some of the children were not allocated after this procedure, but the number remaining amounted to less than 4% of the total number of children living in the region.

To be able to assign children into schools it was necessary to divide the age groups into individual ages. That is, we modeled the exact age of family members, and we did this by uniformly randomly choosing an age within the assigned age group.

Thus for each family in the model population we had the following information: the dwelling identification number, the family identification number, the age groups of the mother and the father (where applicable), and the sex, age group and age for each of the children, if any.

#### 2.1.4. Non-family households

Using the variable "Living alone" in HLA tables (by municipalities), we selected dwellings with individuals in non-family households who either resided alone or did not reside alone but did not constitute a family. These individuals were also assigned to a sex and age group.

To assign those individuals to households who did not reside alone but did not constitute a family, we chose two individuals selected by simple random sampling without replacement to live in each household classified as non-family with more than one individual. Those in the population having this status but not allocated after this procedure were then uniformly randomly assigned to a non-family household with more than one member, subject to municipal totals.

As in the case of family members, we modeled the age of an individual living in a non-family household by randomly selecting an age value within the individual's age group, according to a discrete uniform distribution.

## 2.2. Schools

Peel Public Health obtained data on school catchment areas for the elementary and secondary schools in the local public school board. The GIS databases have the school ID, name and its geographical polygon.

Using the information from the the school geographical polygon, the location of the dwelling, the modeled number of children living in each dwelling and their ages, the children were placed into schools.

## 2.3. Random contact networks

To model the population in the Region of Peel, we considered three basic kinds of subnetworks: "dwellings", "schools" and "general adult", and we defined some general connectivity structures or patterns within subnetworks using random networks.

In the specific case of the "dwelling" subnetworks, we assumed that individuals living in a dwelling have all contact to each other. Then in the file edges.txt we included the edge  $(v_1, v_2)$  if individuals  $v_1$  and  $v_2$  live in the same dwelling.

To model the interactions of in-school children and of outside-dwelling adults subnetworks, we used random networks. Each of these subnetworks was constructed following a drawn set of degrees from a specified degree distribution. Some of the distributions that have been implemented in the program are: binomial, Poisson, zero-truncated Poisson, geometric, polylogarithmic.

Since the subnetworks have finite numbers of vertices, the generated random networks will not follow the specified

distribution exactly; however, the actual distribution would converge to the specified theoretical distribution as the number of vertices increases.

In the case of the networks classified as “schools”, each of them was constructed with the same degree distribution family (such as the Poisson distribution with the mean  $\lambda$ ). In turn, the “general adult” network that describes the relationships between individuals who do not attend school (but interact in other settings), can be modeled with a highly heterogeneous network such as the polylogarithmic distribution.

The selection of the degree distributions and related parameters is not necessarily linked to statistical estimation from data. The degree distributions are chosen to reflect a social interaction structure within the subgroups (which will become subnetworks) that have been already constructed using demographical data. In particular, a user is offered an option to model the “family” and “school” subnetworks as fully connected or Poisson degree distributed. For the “adult” subnetwork we find it more realistic to use a power-law network, i.e. following a polylogarithmic degree distribution. Indeed, the “adult” subnetwork actually represents the combination of multiple subnetworks and, hence, is expected to be very heterogeneous, as reflected in previous studies on power law tails in adult subnetworks [28,29,3]. The selection of the degree distribution parameters is completely arbitrary. However, if external information is available, we can estimate the parameters of some subnetworks and then utilize them for constructing social interaction within these subpopulations.

For instance, the “general adult” network is selected to be heterogeneous (polylogarithmic(1,10)) with mean of 4.04 contacts. This is the subnetwork with the highest order and it takes approximately 7 h to be constructed using a 2.2 GHz computer with 8 GB in RAM. In contrast, it takes less than 20 seconds to construct each of the “school” networks.

Once all the subnetworks were obtained then the new edges were added to the file `edges.txt`, in which we also specified the subnetwork in which each edge occurs.

The finally obtained files `edges.txt` and `indexes.txt` have sizes of 60 and 20 MB, respectively and can be used to simulate multiple outbreaks with different scenarios.

**Remark.** While the currently utilized network generation algorithm requires a longer time than some other algorithms, our primary goal is to construct a random network for the overall population that could allow for a uniformly random selection from the networks that meet a pre-specified degree distribution sequence. Constructing a simple network from a given degree sequence is a classical problem in graph theory and computer science (see, for example, overviews by [33,36]). However, generating a network uniformly at random from all the possible networks with that specific degree sequence is an open problem. Some algorithms include a two-stage procedure of constructing a network and then modifying it sequentially through a Markov chain simulation process [1,33,14]. Although the well known Havel-Hakimi algorithm [16,15] yields an existence-constructive result for networks with a given degree sequence, the topologies of the obtained graphs using this algorithm exhibit very *dense cores*, failing in creating uniformly random graphs among those which meet the degree specifications. The algorithm that we currently employ for SIMID is the Molloy-Reed method [34]. In

the first step of the network construction, the edge  $(i;j)$  is selected with probability proportional to the product of the degrees  $d(i)$  and  $d(j)$ . After an edge is selected, the degree of the vertices is updated to consider the connections that each vertex still can use, called the residual degree. In the subsequent steps, the selection of the edges is performed with probabilities proportional to the product of the residual degrees. This network generation algorithm allows for any pre-specified degree sequence for the network or sequences drawn from the following probability distributions: Poisson, zero-truncated Poisson, geometric, zero-truncated geometric, negative binomial, polylogarithmic and logarithmic. The algorithm also allows the option “full” that constructs a complete network, i.e. where any two vertices in the network are connected. The simulation program developed is written in the R language and an earlier version was available as the R package “InfNet”.

While at the stage of developing the SIMID tool, the authors were unaware of the EpiFire software [17] and its capabilities to save and export the network to be used by other programs, we plan to incorporate EpiFire into the next version of SIMID.

### 3. Methods: simulation of spread of infection

In this section we describe the algorithm for statistical scenario generation and its key options and characteristics. In particular, note that the simulation algorithms for the infection process account for different contact structures among individuals in a population, while the infection process evolves as a stochastic process based on the transmission rates (or probabilities) between infective and susceptible individuals.

In the Appendix we list the parameters that can be used in the simulation of the infectious process, given a population and its specified network of contacts (`indexes.txt` and `edges.txt`). In this section we describe SIMID’s general characteristics and the scenarios it includes.

In our SIMID tool, we implement two compartmental models: SIR and SEIR. These two models differ according to the status of an individual when he is infected during the outbreak. In the case of SIR, the individual is susceptible to the infectious agent, then is infected and able to transmit the disease (*infective*), and then *recovers* (or is *removed*). This last stage assumes that after infection, the person cannot be re-infected during the outbreak, due to either gaining immunity or death.

In the case of the SEIR model, the susceptible individual after infection acquires the status of *exposed*, which indicates that he has been infected but he is not yet infectious. Then after a period of time he becomes infectious, and after a further time he is removed.

The period of time that an individual can stay in the “E” and “I” stages can be fixed or modeled as random variables. In the simulation program we can incorporate various probability distributions for the lengths of the E and I stages.

In the current version of the program we model the removal of individuals not only due to the termination of the infectious period, but also by some control measures such as vaccination and isolation.

### 3.1. Infection process

In order to decrease the running time of the simulation program, we consider an algorithm that describes the infectious process by days without imposing a restriction that distributions for the exposed and infectious periods are discrete. The individual infection process can be described as follows:

- **S stage.** If on day  $i$  vertex  $v$  is infected, the infectious hour is recorded as  $24i - 12$ . The initial cases or “patient zeros” are the only ones that are considered to be infected at hour 0.
- **E stage.** If the epidemic model is SEIR, after being infected the individual remains infected (and not infectious) for a period of time ( $h_e$ ). The length of this stage is defined by the exposed (latent) distribution and its parameters.
- **I stage.** Once the exposed period is completed (at time  $24i - 12 + h_e$ ), the individual can start passing the infection to those individuals with whom he is connected. The infection transmissibility is driven by three parameters, namely, BETA1, BETA2 and t.rates. The value of BETA1 is only considered as a rate when t.rates=TRUE. In this case the transmission to a connected susceptible individual (during the infectious period) follows a Poisson process with rate BETA1. If t.rates=FALSE, then BETA1 represents the probability that infectious individual  $v$  passes the infection to a connected susceptible individual during a pre-specified unit of time (e.g. hour). In turn, BETA2 is the parameter for the level of susceptibility, relative to a “normal” level. For example, if an infectious individual transmits the agent with a parameter BETA1 and is connected to a susceptible individual whose susceptibility is only half the susceptibility of “normal” susceptible, then BETA2 = 1/2 for that individual. Hence, BETA2 is viewed as an individual characteristic of susceptibility. BETA1 and BETA2 can be different for the age groups delineated by the variables breaks1 and breaks2. More details of their default values and how to specify them are in the Appendix.
- **R stage.** The duration of the infectious period ( $h_i$ ) is defined by the infectious distribution and its parameters. After this time, the individual is considered as removed from the infectious system (at time  $24i - 12 + h_e + h_i$ ).

The units for the lengths of exposed and infectious period are in hours, and these lengths can be discrete or continuous random variables, such as constant, normal, lognormal, exponential, Poisson (see the Appendix). However, due to the fact that the simulation algorithms evolve by days, identifying each time individuals that are infectious or are to become infectious during the next day, the latent period must be longer than 12 h. This assumption is made in order to avoid having the infectious period start the same day as the infection of the newly infected person.

The transmissions between edges are considered to be independent and are obtained for each day  $i$  and for each connection between a susceptible individual  $v$  and an infectious (or infected and about-to-become infectious) individual  $w$ .

If BETA1 is a rate this probability is equal to

$$P(w \text{ infects } v \text{ during day } i) = 1 - \exp^{-\text{BETA1}_w \times \text{BETA2}_v \times T_{vw} \times I_{w_i} / 24}$$

and if BETA1 is a probability, then the probability of transmission is

$$P(w \text{ infects } v \text{ during day } i) = 1 - (1 - \text{BETA1}_w \times \text{BETA2}_v)^{T_{vw} \times I_{w_i} / 24},$$

where  $\text{BETA1}_w$  and  $\text{BETA2}_v$  are the transmission and susceptibility parameters of individuals  $w$  and  $v$ , respectively;  $T_{vw}$  is the number of hours the edge between  $v$  and  $w$  is active each day (for details see the variable time.subnetwork in the Appendix), and  $I_{w_i}$  is the number of hours the infectious individual is able to transmit the infectious agent during day  $i$ .

**Remark.** Note that both the “percolation model” and the “chain binomial model” considered in the EpiFire package (see [17]) can be viewed as subcases of the SIMID model. In particular, the EpiFire “percolation model” represents the Reed-Frost model evolving in the network of contacts, where the latent period equals to 1, the infectious period is a fixed  $\epsilon$ ,  $\epsilon \ll 1$ , and the fixed probability of transmission (to the connected susceptible individuals) is  $T$ . In turn, the EpiFire “chain binomial model” represents a discrete-time epidemic model that incorporates in infectious period (variable called gamma). Then, the resulting epidemic model is the same as the “percolation model”, with the exception that the infected individual (vertex) is not removed immediately after it is able to transmit the infectious agent to its neighbors, but instead remains infective for a fixed period gamma. Note that the model in the SIMID approach is more general than the “percolation model” and the “chain binomial model”. While the SIMID network model also evolves in discrete time (days), it allows both the latent and infectious periods to be random variables, rather than only pre-defined constants, and these random variables for latent and infectious periods can follow different probability distributions for various individual (vertex) age-groups. The only restriction on the probability distribution in the SIMID network model is that the latent period has to be larger than 12 h, so that a recently infected individual does not become infective during the same day. This restriction is imposed in order to keep the computer algorithm simple. From a practical perspective of modeling spread of infectious diseases, this restriction is relatively minor, since most infections with an identified latent period typically exhibit latent periods longer than 1 day.

Each outbreak simulation that employs previously obtained indexes.txt and edges.txt files, takes between 15 and 120 minutes to be completed, using a 2.2 GHz computer with 8 GB in RAM. The difference in time between simulations is directly linked to an outbreak size. When the parameters are defined under an outbreak threshold, all outbreak sizes remain small and the simulation time is approximately 15–30 minutes.

### 3.2. Interventions

Three basic intervention processes are incorporated into the program. The first is vaccination, that can also be used to model the natural immunity of individuals. The other two interventions are the isolation of individuals within specified subnetworks and the elimination of subnetworks during the outbreaks.

The function *vaccine* is used to implement the vaccination as described in the next two sections. This control measure can be carried out at time 0, or in the same way as the subnetwork isolation and elimination, can be implemented during the outbreak. The parameters of vaccination are the parameters for selection and new transmissibilities.

There are three different criteria indicating when a control measure has to be implemented. They are selected through the variables *crit* and *param.crit*, and incorporated in the function *criteria.pause*. When the variable *crit* has the value 1, the intervention is instantaneously done between day (*param.crit*-1) and *param.crit*.

When *crit* is equal to 2, the intervention is done as soon as the total number of infected, including “patient zero(s)” and starting at time 0, reaches the critical specified value *param.crit*.

When *crit* is 3, the intervention is done as soon as the total number of new infected reaches the pre-specified critical value of *n.infectious* in the last consecutive *n.days* days.

For example if *crit* = 1 and *param.crit* = 3, then the intervention is not yet carried out at the end of day 2 but considered fully implemented at the first hour of day 3.

If *crit* = 2, *param.crit* = 8, we have one patient zero and the number of new cases for days 1, 2, 3, and 4 are: 2,4,0,2, respectively, then the intervention is implemented at the first hour of day 5.

If *crit* = 3, *param.crit* = 5, *n.days* = 2, we have one patient zero and the number of new cases for days 1, ..., 6 are: 2,1,0,2,3,1, then the control measure is executed at the first hour of day 6.

### 3.3. Vaccination

#### 3.3.1. Parameters for selection

We can define two possible scenarios when selecting individuals to be vaccinated: in the first one we vaccinate both susceptible and infected individuals, and in the second one, we vaccinate only susceptible individuals. The first case applies to infections that cannot be identified until the onset of symptoms. Even though the latent period does not always coincide with the asymptomatic period, this case is introduced as an approximation.

The simulation assumes that the vaccine does not have any effect on already infected individuals.

The second scenario (when variable *s.only* = TRUE) is less realistic since it assumes that we can differentiate between susceptible and infected (and asymptomatic) individuals.

The number of selected individuals for vaccinations can be set as a fraction (number between 0 and 1) of candidates to vaccinate or a specific number. The targets for selection of individuals can be set for age groups (described by *break.vac*) and if the number is larger than the total number of vaccine-candidate individuals in an age group, the total of susceptible and infected (*s.only* = F) or susceptible (*s.only* = T) is vaccinated.

#### 3.3.2. Parameters to describe new transmissibilities

For this version of the program we consider that the vaccine does not modify the latent or the infectious period, but it modifies the transmission parameters BETA1 and BETA2. Then the transmission parameters for vaccinated individuals

become equal to those specified by *new.Beta1* and *new.Beta2*. Like BETA1 and BETA2, *new.Beta1* and *new.Beta2* can be different for specified age groups, delineated by *new.breaks1* and *new.breaks2*. These new age groups are not required to have any particular relationship with their unvaccinated transmissibility counterparts *breaks1* and *breaks2*.

When the vaccine is 100% effective for protecting individuals from infections, we can set *new.Beta2* = 0 and *new.breaks2* equal to the vector (0,100), to indicate that *new.Beta2* is defined for individual between 0 and 100 years old.

If the vaccine is not 100% effective but reduces the acquisition to  $q \times 100\%$  then *new.Beta2* = *q*. In this case, we may be interested in defining a new value for *new.Beta1* since the vaccine may also reduce the transmissibility of those individuals that were infected after vaccination.

### 3.4. Subnetwork isolation and elimination

The intervention related to the containment of individuals within specific subnetworks is included in the program with the function called *isolate.subnetwork*. In terms of graphs, this action is equivalent to identifying all the individuals with contacts within the subnetworks to isolate, and eliminating all the edges they have with these subnetworks. The main parameter of this function is a list with the subnetwork ID's to isolate.

In addition, the elimination of subnetworks can correspond to actions such as school or workplace closures. This measure is implemented in the function *eliminate.subnetworks* and eliminates all the edges connecting within specific subnetworks. Its parameter is a vector with the subnetwork ID's to eliminate.

The functions *isolate.subnetwork* and *eliminate.subnetworks* are included in the function *main.epid.sim* but only one control measure during an outbreak is allowed in the current version. Inclusion of several interventions can be easily added to *eliminate.subnetworks*.

## 4. Technical specifications and architecture

The SIMID tool is developed using the following components, and made accessible through a web interface. The system components are broken down into the following, and each is explained below:

### 1 Internet Information Services (IIS) and Secure Socket Layer (SSL)

IIS and SSL are the foundation and entry point into the disease simulation website. A Windows server furnished by the Ontario Agency for Health Protection and Promotion (OAHPP) running IIS 6 provided the building blocks for all application interfaces. Using the simulation manager, power users are able to set parameters and run simulations to be consumed by users who have viewing or read only access to simulations. Other interfaces such as user security and management, user login and simulation manager run under the IIS framework. All other non-GIS interfaces are deployed under this framework. SSL and IIS are used to deploy testing, staging and production disease simulation websites. The interface for this server uses 128 bit encryption with a SSL. All internal and external transmissions are encrypted using this protocol.



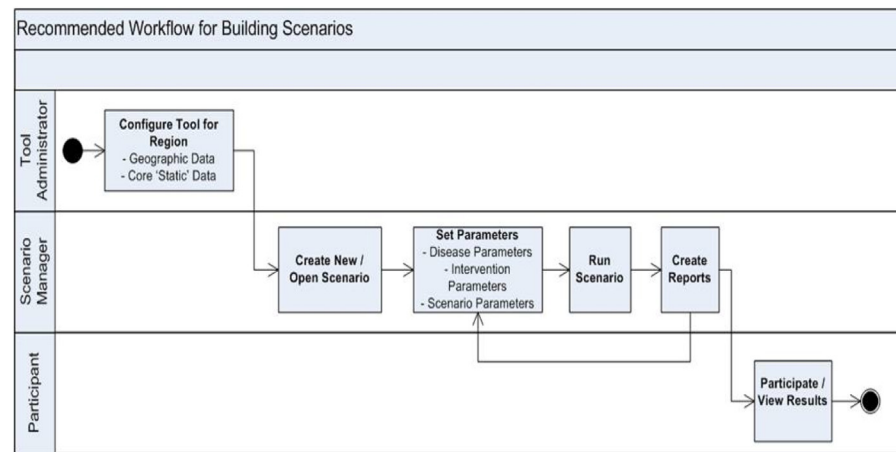


Fig. 1 – The SIMID flow chart.

## 2 ArcGIS Server 9.3 Enterprise Standard Service Pack 1

ArcGIS Server 9.3 Enterprise Standard Service Pack 1 technology supports the GIS tier of the disease simulation system. A Windows server furnished by the OAHPP running ArcGIS Server 9.3 provided the building blocks for all GIS application interfaces. The interface for this server uses 128 bit encryption with a SSL. All internal and external transmissions are encrypted using this protocol. Map services are configured using this technology and deployed as interfaces to be consumed and integrated with the Silverlight interface of the disease simulation tool. ArcGIS Server technology also handles the load balancing for simulation map request. All GIS based data are stored in a local file.

## 3 ESRI Toolkit Silverlight 3

All GIS-based web interfaces are integrated into the IIS website using ESRI Toolkit Silverlight 3. This interface provides the tools to view time series GIS data over geographic areas within the Peel Public Health catchment area and to comparatively view temporal data throughout the observation period of a simulation.

## 4 SQL Server 2005

SQL Server 2005 provides the business data repository for SIMID. A Windows server furnished by the OAHPP was allocated for this software. The results of simulation-based R software data are aggregated and stored in the SIMID database. All identifiable information stored in the database is encrypted at the field level. The simulation manager makes exclusive use of stored procedures, tables and views to aggregate the simulation data so that it could be spatially joined with GIS data and viewed on a map. This information is cartographically displayed on a map with other layers from the ArcGIS Server 9.3 service.

## 5 R Analysis Software

R software is used as the engine for executing simulations for random networks within SIMID. A Linux-based server was allocated to run this software by the OAHPP. R Software setup and configuration is handled by the University of Waterloo<sup>3</sup> and OAHPP. Data outputs from R are used by

ArcGIS Server and the simulation manager and published to SQL Server.

## 5. SIMID user interface

In this Section we discuss the SIMID application from a user's point of view and present a sample run.

Fig. 1 represents the recommended flow chart for generating scenarios for infectious disease outbreaks, using the SIMID tool whose execution can be summarized by the following algorithm:

- 1 'Scenario Manager' parameterizes and tests a simulation in DRAFT.
- 2 Scenario Manager chooses to RUN a simulation.
- 3 Job Service tells R software that a simulation is 'waiting' and executes the simulation in R.
- 4 When a simulation is completed in R, data output is created.
- 5 Job Service 'listens' to SIMID and determines whether the simulation is completed.
- 6 Once the simulation is completed, Job Service takes the completed output generated by R and imports data into SIMID.
- 7 Simulation data are used by SIMID to generate map outputs and other simulation statistics.

Now we provide a sample run of the SIMID tool as outlined below:

- First, the user logs in, which identifies his/her role ('Scenario Manager' or 'Participant') (see Fig. 2).
- The Home screen (see the left panel of Fig. 3) shows a list of simulations and their status, while the right panel of Fig. 3 presents information on 'running' simulations that are being processed by the 'R' server.
- As shown by Fig. hyperlinkstatus4, the user can overview the status of the user's simulations. Fig. 5 shows the lifecycle of a simulation, from DRAFT, to RUNNING, to FINALIZED, to ARCHIVED status.
- High level information about a user's DRAFT simulations can be found in Fig. 6.

<sup>3</sup> The R routine is available upon request from Lilia Leticia Ramirez Ramirez.

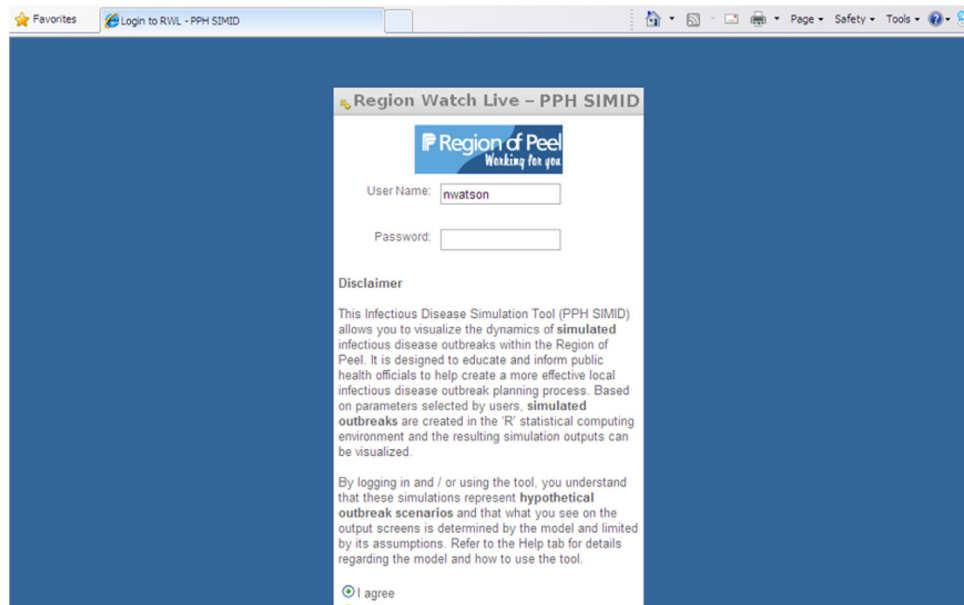


Fig. 2 – Step 1.

- Fig. 7 shows how the user can customize various simulation parameters.
- Finally, spatial outputs for an outbreak scenario and progression can be presented on a map of the region along with the output for other important outbreak characteristics, and a time-slider enables spatiotemporal visualization of the simulated outbreak (see Fig. 8).

Hence, with the help of the SIMID tool, complex disease network models developed in the ‘R’ language can be parameterized, run and visualized by field personnel who are designated ‘Scenario Managers’, while Collaborative Modeling Scenario Managers can explore simulations,

change parameters and then share outputs with designated ‘Participants’. Simulations can be developed to support ‘table-top’ exercises with high-fidelity scenarios, which cultivates a ‘Learning Organization’, creates intra/inter-organizational knowledge by testing assumptions and exploring “what if” scenarios, e.g. contrasting simulated interventions, evaluating response times, and quantifying costs.

## 6. Discussion

The SIMID program, based on modern random network methodology and integrating complex epidemiological modelling in R with space–time visualization in GIS, provides

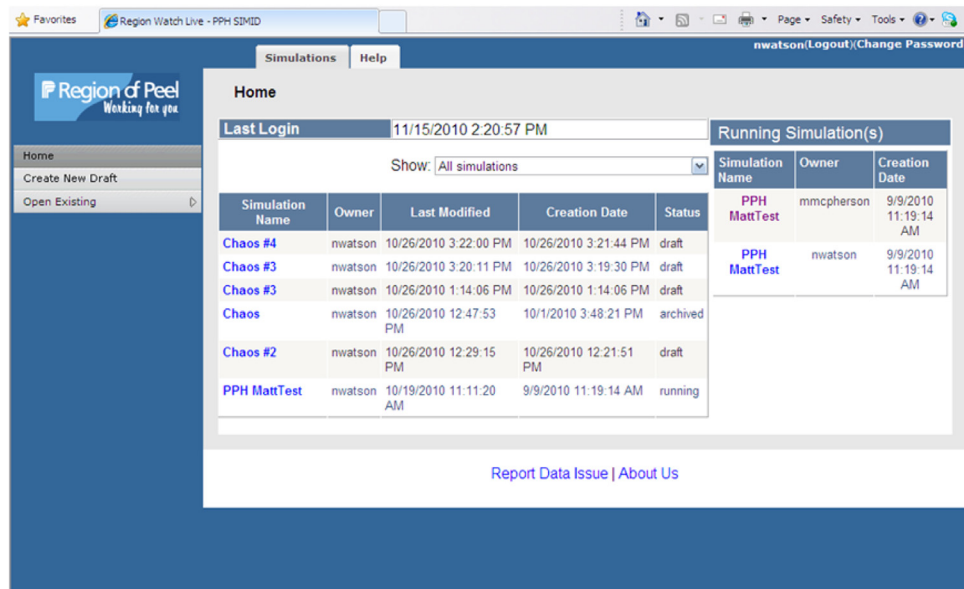


Fig. 3 – Step 2.

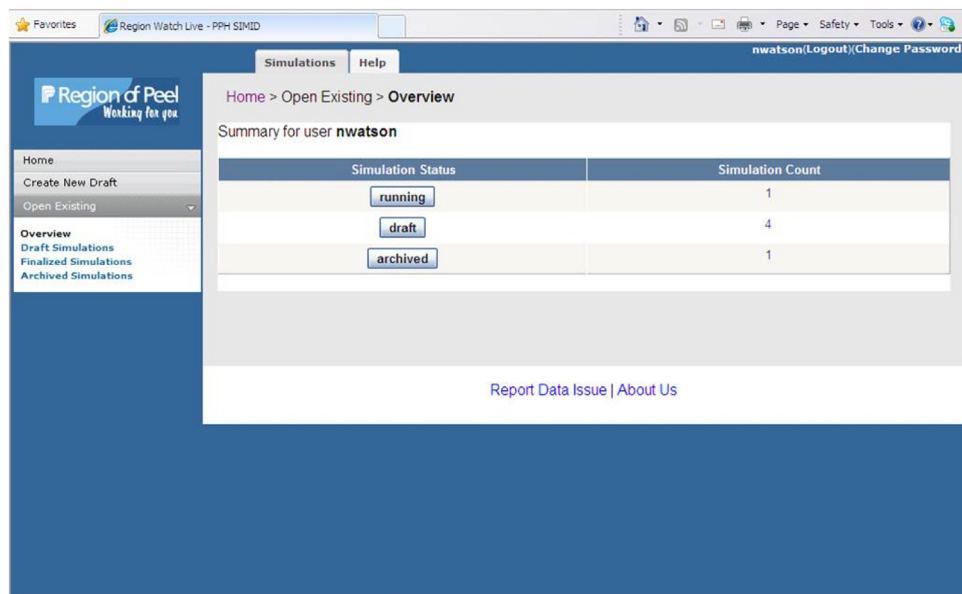


Fig. 4 – Step 3.

### Simulation Lifecycle

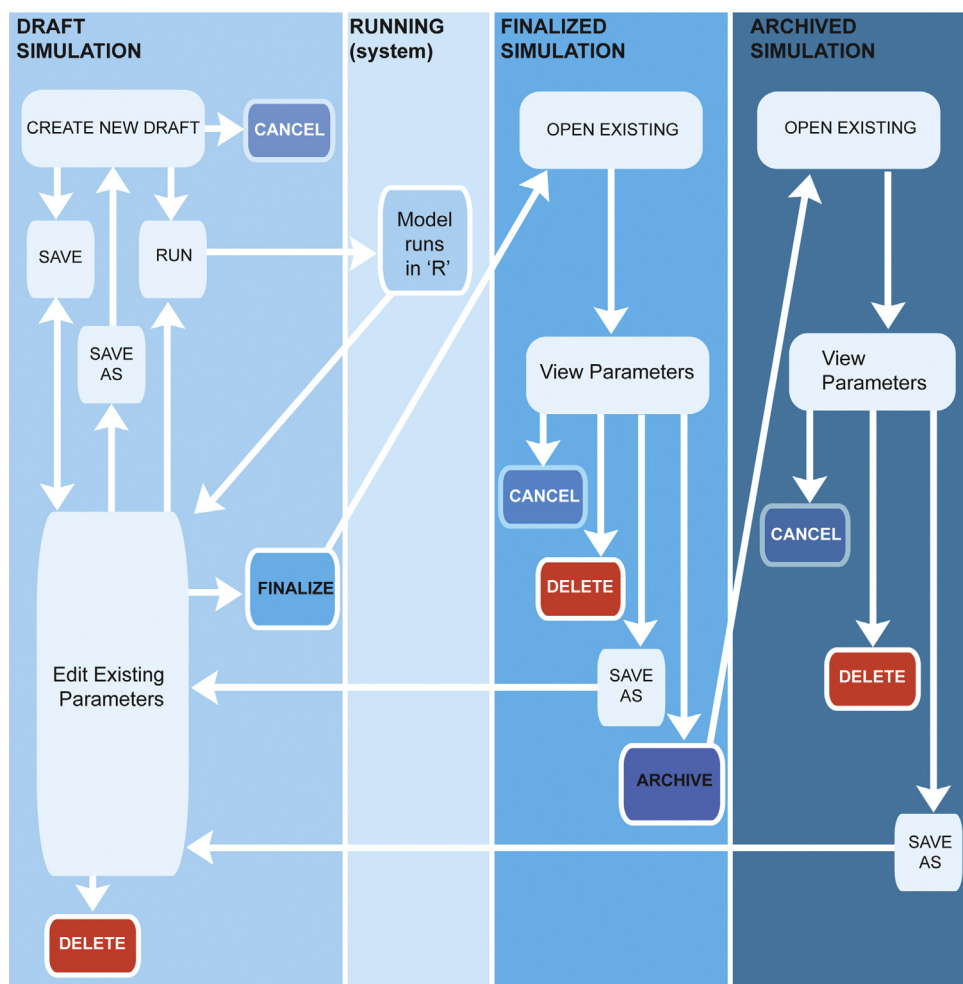


Fig. 5 – Lifecycle of a simulation.

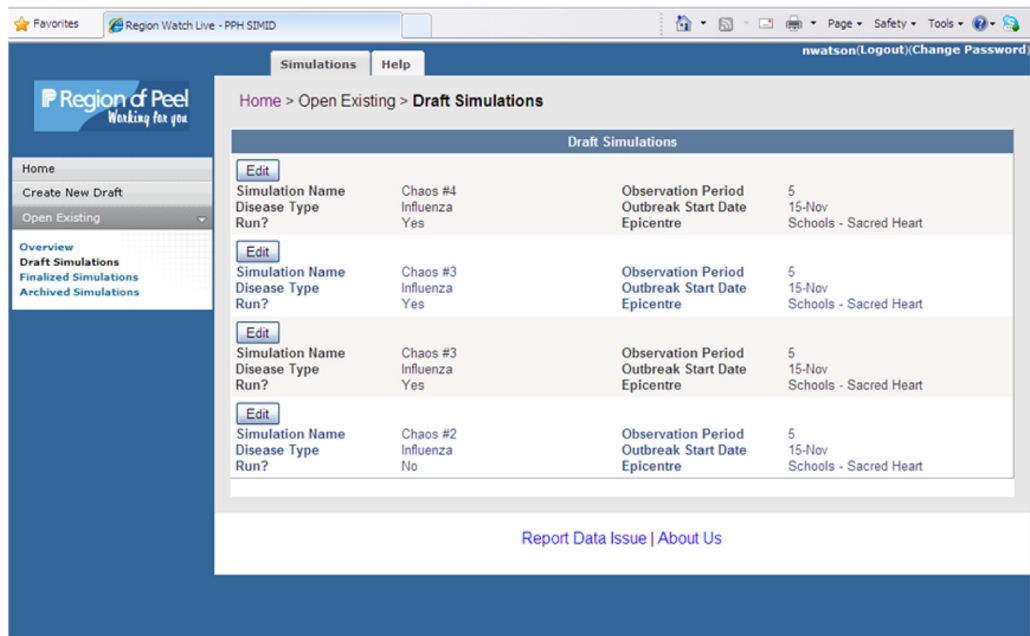


Fig. 6 – Information on draft simulations.

health professionals and decision makers with more effective visualization and assessment of the dynamics of the infectious outbreak over time within their own local communities. Hence, SIMID can be employed for both training and planning purposes, and in particular, to simulate readiness and response, evaluate response times, quantify costs, and ultimately assess the impact of different public health interventions.

While the current application of SIMID focuses on influenza outbreaks, including both pH1N1 and seasonal influenza in the Region of Peel, Ontario, Canada, we can readily extend this

analysis to any region in Canada or other country, using the relevant demographic or census data on families and information on schools.

Moreover, the SIMID tool is simple and flexible in structure and, conditioned on the data availability, can be further advanced to include various subnetworks, reflecting contacts within hospitals, long term care facilities and workplaces as well as taking into account travel patterns to and from these places. Also, the system design paradigm and technology platform are applicable to a wide spectrum of communicable diseases and health-related urgent response events,

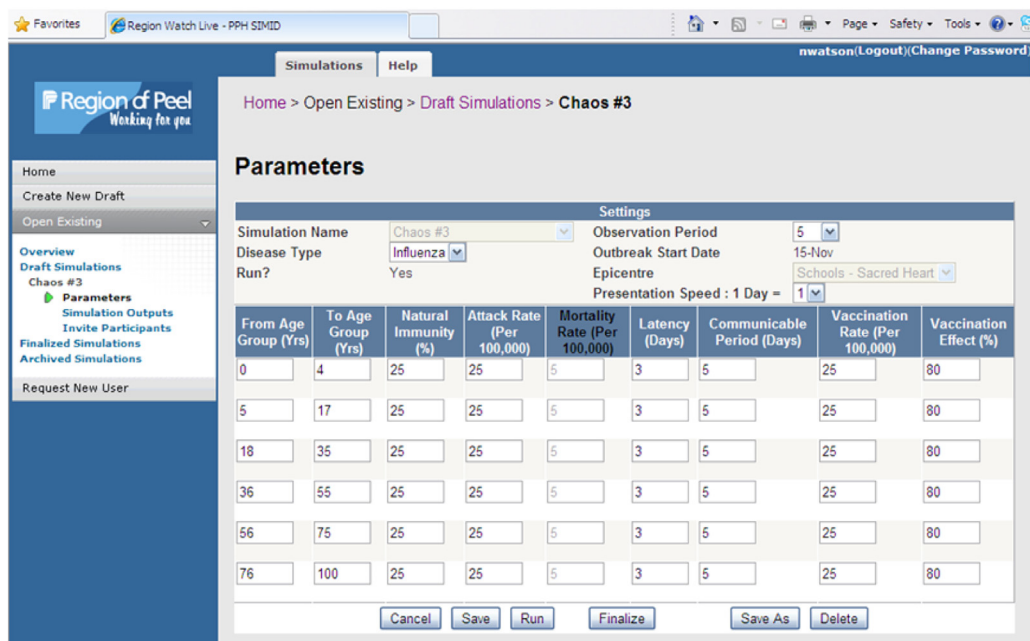


Fig. 7 – Customization of model parameters.



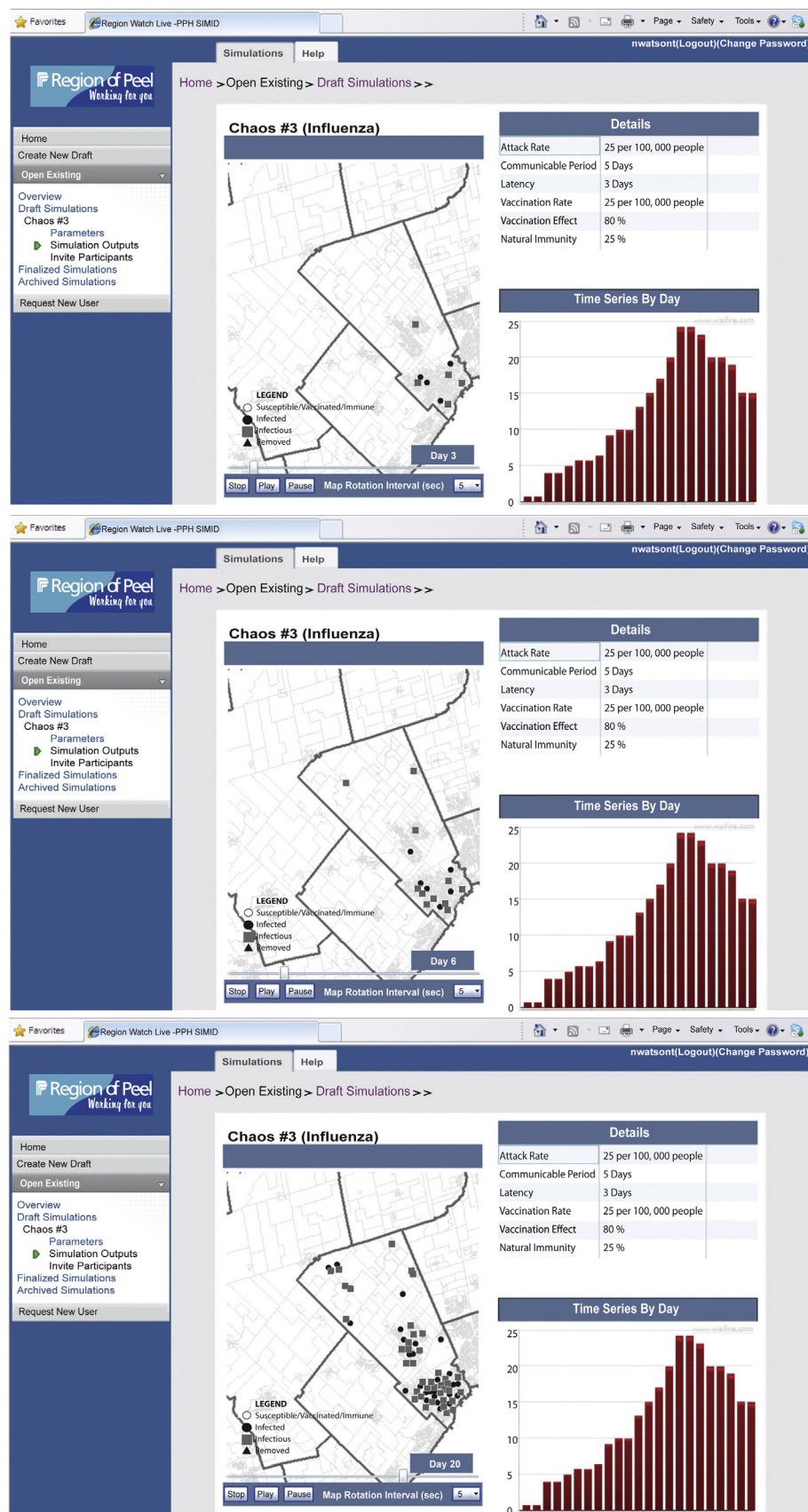


Fig. 8 – Visualization of the outbreak scenarios and spatial progression.

and are extendible to incorporate the effects of weather-related, chemical, biological, radiological, nuclear and other environmental health factors. In turn, the network model can be further advanced by generalizing transmissibility and susceptibility random variables with individual probability distributions such that the individual distribution can be also governed by environmental heterogeneities. Moreover, we may also integrate our random network epidemiological model with mapping procedures from geostatistics, in order to advance visualization and space–time downscaling of infectious dynamics.

## Acknowledgments

The project was supported by GeoConnections. In addition, the research of Gel and Thompson was supported by grants from the Natural Science and Engineering Research Council (NSERC) of Canada, and the research of Ramirez Ramirez was supported by MITACS Accelerate internships and the Fields postdoctoral scholarship. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: <http://www.sharcnet.ca>) and the support of Asociación Mexicana de Cultura A.C. We would like to thank Vyacheslav Lyubchich, University of Waterloo, for assistance with simulations at the revision stage.

## Appendix A.

The function for simulation of epidemic dynamic (with at most one control measure and immunity) is *main.epid.sim*. Its parameters are:

- *indexes*. This is a text variable with the filename of where the *indexes.txt* file is stored. It includes its path if necessary.
- *fEdges*. This is a text variable with the filename of where the *edges.txt* file is stored. It includes its path if necessary.
- *t.rates*. This variable indicates whether the parameter BETA1 is a rate or a probability of transmission per unit of time. Its default value is TRUE, indicating BETA1 is a rate.
- *id.infected*. This variable is a vector with the individual ID's that are infected at time 0. Its default value is NULL, to indicate that no individual is initially infected.
- *id.infective*. Like *id.infected*, this is the vector with the individual ID's of those who are infective at time 0. Its default value is NULL.
- *ini.infected*. This is a two-column matrix that indicates the number of initially infected individuals in the specified subnetwork. Its format is: 1-subnetwork ID, 2-number of cases. The infected individuals are randomly selected within each subnetwork. Its default value is NULL.
- *ini.infective*. This is a two-column matrix that indicates the number of initially infective individuals in each specified subnetwork. Its default value is NULL.
- *BETA1*. This is the rate (or probability) of infection per hour. It is a scalar or a vector if different values of transmissibility are set according to age groups.
- *breaks1*. This encodes the age groups used for different values of BETA1. Then if BETA1 is a scalar (*breaks1* is the vector

(0,100) -default value-) and if BETA1 is a vector of length  $n$  then *breaks1* must be a vector of length  $n+1$ . The first value of this vector must be 0 and the last must be 100. For example: *BETA1* = (0.1,0.15,0.20), *breaks1* = (0,1,70,100).

- *BETA2*. This parameter indicates the susceptibility and like BETA1, it can be a scalar or a vector. Its default value is 1 that corresponds to the case when all individuals are equally susceptible to infection.
- *breaks2*. The age groups for BETA2. Its default value is (0,100).
- *distrib.lat*. This is the name of the distribution for the latent period. Its default value is “norm” (Gaussian distribution), but it can be set to be “exp” (Exponential), “lnorm” (Lognormal) or “mass” (Point or degenerated distribution).
- *LAMBDA*. This is the parameter for the latent period. It can be a scalar for some distribution functions or a vector for some others. Its default value is NULL. In the same way as for BETA1 and BETA2 we can set different parameter values by specific age groups.
- *breaks.lambda*. The age groups for LAMBDA. Its default value is (0,100).
- *distrib.inf*. This is the distribution name for the infectious period.
- *GAMMA*. This is the parameter for the infectious period and like LAMBDA can be different according to age group.
- *breaks.gamma*. The age groups for GAMMA. Its default value is (0,100).
- *time.subnetwork*. Indicates the number of hours edges are active during the day. Its default value is 24 to indicate that all connections are active 24 h, but this value can be set differently according to the subnetworks the edges belong to. An edge is classified as belonging to a subnetwork  $v$  if both endpoints are in subnetwork  $v$ . This information is in the file *edges.txt*. In *time.subnetwork* different values are set in a matrix where the first column is the subnetwork ID and the second one is the time (in hours). Under the scenario that the infectious period is shorter than 24 h for a particular individual, the time he/she spends connected and infectious is proportional to the number of hours he/she is infectious in that day.
- *seir*. This is a flag variable and is TRUE for a SEIR model and FALSE for a SIR model. Its default value is FALSE.
- *obs.time*. The epidemic simulation period in days. Its default value is  $10^6$ .
- *crit*. This is the criterion as specified in Section 3.2. Its default value is 0 to indicate that no vaccination intervention is done after day 0.
- *param.crit*. The parameter of *crit*.
- *vac0*. This is a flag variable that and it is set TRUE when vaccination prior to outbreaks (or natural immunity) is present. Its default value is FALSE to indicate that all non infected (infective) individuals are susceptible.
- *frac0*. This is the fraction of immune individuals as described in Section 3.3.1. Its default value is NULL.
- *number0*. This is the number of immune individuals. Its default value is NULL.
- *break.vac0*. The age groups for *frac0* or *number0*.
- *vac*. This is a flag variable that is set TRUE when vaccination is implemented during the outbreak. Its default value is FALSE.

- `s.only`. This is a flag variable that indicates if the individuals to be vaccinated are from the population of susceptible or infected individuals (FALSE), or just from the population of susceptible individuals (TRUE -default-).
- `frac`. Similar to `frac0` but related to vaccination. It is the fraction of vaccinated individuals during the outbreak.
- `number`. Similar to `number0`, it is the number of vaccinated individuals during outbreak.
- `break.vac`. Similar to `break.vac0`.
- `new.Beta1`. It is the rate (or probability) of infection per hour for those individuals that have been vaccinated but are infectious either because 1) they were vaccinated when they were already infected or 2) because the vaccine does not always prevent from infection. In the current version both vaccinations (at time 0 and during the outbreak) have similar effects on the transmissibility. This can be easily changed in the program *main.epid.sim*.
- `new.breaks1`. The age groups for `new.Beta1`. Its default value is (0,100).
- `new.Beta2`. Similar to `new.Beta1`.
- `new.breaks2`. The age groups for `new.Beta2`. Its default value is (0,100).
- `elimSubnet`. The flag variable to indicate subnetwork elimination during the outbreak. Its default value is FALSE.
- `closeSubnet`. This is a flag variable to indicate subnetwork isolation. Its default value is FALSE.
- `subnetworks`. The subnetwork' ID's to eliminate (`elimSubnet=TRUE`) or to isolate (`closeSubnet=TRUE`). Its default value is NULL.
- `filenameOut`. The name of the output file that contains the description of the epidemic simulation.

## REFERENCES

- [1] W. Aiello, F.R.K. Chung, L. Lu, A random graph model for power law graphs, *Experimental Mathematics* 10 (2001) 53–66.
- [2] L. Ancel Meyers, Contact network epidemiology: bond percolation applied to infectious disease prediction and control, *Bulletin of the American Mathematical Society* 44 (2007) 63–86.
- [3] L. Ancel Meyers, B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, R.C. Brunham, Network theory and SARS: predicting outbreak diversity, *Journal of Theoretical Biology* 232 (2005) 71–81.
- [4] L. Ancel Meyers, M. Newman, M. Martin, S. Schrang, Applying network theory epidemics: control measures for outbreaks of *Mycoplama pneumoniae*, *Emerging Infectious Diseases* 9 (2003) 204–210.
- [5] H. Andersson, Limit theorems for a random graph epidemic model, *Annals of Applied Probability* 8 (1998) 1331–1349.
- [6] H. Chen, D. Zeng, P. Yang, *Infectious Disease Informatics: Syndromic Surveillance for Public Health and BioDefense*, Springer, New York, 2009.
- [7] F. Brauer, P. Van den Driessche, J. Wu, L.J.S. Allen (Eds.), *Mathematical Epidemiology*, Springer-Verlag, Berlin, 2008.
- [8] E. de Villa, M. McPherson, L.L. Ramirez Ramirez, 1; SIMID: SIMulation of Infectious Disease, The Ontario Public Health Convention (TOPHC), April 5–8, 2011, Toronto, Canada.
- [9] U.F. Ekpo, C.F. Mafiana, C.O. Adeofun, A.R. Solarin, A.B. Idowu, Geographical information system and predictive risk maps of urinary schistosomiasis in Ogun State, Nigeria, *BMC Infectious Diseases* 874 (2008), OpenURL.
- [10] J.M. Epstein, Modelling to contain pandemics, *Nature* 460 (2009) 687.
- [11] J.M. Epstein, *Generative Social Science: Studies in Agent-based Computational Modeling*, Princeton University Press, Princeton, NJ, 2006.
- [12] N.M. Ferguson, D.A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn S., D.S. Burke, Strategies for containing an emerging influenza pandemic in Southeast Asia, *Nature* 437 (8) (2005) 209–214.
- [13] J. Glaz, J. Naus, S. Wallenstein, *Scan Statistics*, Springer-Verlag, New York, 2001.
- [14] C. Gkantsidis, M. Mihail, E. Zegura, The Markov chain simulation method for generating connected power law random graphs., in: *Proc. 5th Workshop on Algorithm Engineering and Experiments (ALENEX)*, SIAM, 2003, pp. 16–25.
- [15] S. Hakimi, On the realizability of a set of integers as degrees of the vertices of a graph, *SIAM Journal on Applied Mathematics (SIAP)* 11 (1) (1963) 135–147.
- [16] V. Havel, A remark on the existence of finite graphs, *Časopis pro pěstování matematiky* 80 (1955) 477–480.
- [17] J.T. Hladish, E. Melamud, L.A. Barrera, A. Galvani, L. Ancel Meyers, EpiFire: an open source C++ library and application for contact network epidemiology, *BMC Bioinformatics* 13 (2012) 76.
- [18] D.E. Holms, L. Jain (Eds.), *Innovations in Bayesian Networks: Theory and Applications*, Springer-Verlag, Berlin, 2008.
- [19] M.J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton, NJ, 2007.
- [20] A. Kim, A. Martinez, J. Klausner, J. Goldenson, C. Kent, S. Liska, W. McFarland, Use of sentinel surveillance and geographic information systems to monitor trends in HIV prevalence, incidence, and related risk behavior among women undergoing syphilis screening in a jail setting, *Journal of Urban Health* 86 (1) (2008) 79–92.
- [21] P.R. Kleindorfer, Y. Wind, R.E. Gunther (Eds.), *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*, Pearson Prentice Hall, Upper Saddle River, NJ, 2009.
- [22] K. Kleinman, R. Lazarus, R. Platt, A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism, *American Journal of Epidemiology* 159 (2004) 217–224.
- [23] M. Kulldorff, Prospective time periodic geographical disease surveillance using a scan statistic, *Journal of the Royal Statistical Society, Series A* 164 (1) (2001) 61–72.
- [24] M. Kulldorff, R. Heffernan, J. Hartman, R. Assuncao, F. Mostashari, A space-time permutation scan statistic for disease outbreak detection, *PLoS Medicine* 2 (3) (2005) e59.
- [25] A.B. Lawson, K. Kleinman, *Spatial and Syndromic Surveillance for Public Health*, John Wiley, West Sussex, 2005.
- [26] A.B. Lawson, *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, CRC Press, Boca Raton, FL, 2009.
- [27] L.M. Lee, S.B. Thacker, M.E. St Louis, S.M. Teutsch (Eds.), *Principles and Practice of Public Health Surveillance*, Oxford University Press, Oxford, 2010.
- [28] F. Liljeros, C.R. Edling, L.A. Amaral, H.E. Stanley, Y. Aberg, The web of human sexual contacts, *Nature* 411 (2001) 907–908.
- [29] F. Liljeros, C.R. Edling, L.A. Amaral, Sexual networks: implications for the transmission of sexually transmitted infections, *Microbes and Infection* 5 (2003) 189–196.
- [30] J.S. Lombardo, D.L. Buckeridge, *Disease Surveillance: A Public Health Informatics Approach*, Wiley, 2007.

- [31] I.M. Longini, M.E. Halloran, A. Nizam, Y. Yang, Containing pandemic influenza with antiviral agents, *American Journal of Epidemiology* 159 (7) (2004) 623–633.
- [32] I.M. Longini, S.K. Seaholm, E. Ackerman, J. Koopman, A.S. Monto, Simulation studies of influenza epidemics: assessment of parameter estimation and sensitivity, *International Journal of Epidemiology* 13 (4) (1984) 496–501.
- [33] M. Mihail, N. Vishnoi,;1; On generating graphs with prescribed vertex degrees for complex network modeling. Position Paper, Approx. and Randomized Algorithms for Communication Networks (ARACNE), 2002.
- [34] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, *Random Structures and Algorithms* 6 (2-3) (1995) 161–180.
- [35] M.E.J. Newman, The spread of epidemic disease on networks, *Physical Review E* 66 (2002) 016128.
- [36] M.E.J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, 2010.
- [37] D.U. Pfeiffer, T.P. Robinson, M. Stevenson, K.B. Stevens, D.J. Rogers, A.C.A. Clements, *Spatial Analysis in Epidemiology*, Oxford University Press, Oxford, 2008.
- [38] L.L. Ramírez-Ramírez, On the dynamics of infectious diseases in non-homogeneous populations, PhD dissertation, University of Waterloo, 2008.
- [39] L.L. Ramírez-Ramírez, M.E. Thompson, Applications of the variance of final outbreak size for disease spreading in networks, To appear in *Methodology and Computing in Applied Probability* (2013), <http://dx.doi.org/10.1007/s11009-013-9325-z>.
- [40] L.L. Ramírez-Ramírez, M.E. Thompson, Infectious diseases in hierarchical networks, Working Paper, University of Waterloo, 2013.
- [41] C. Robertson, T.A. Nelson, Review of software for space–time disease surveillance, *International Journal of Health Geographics* (2010) 9–16.
- [42] C. Robertson, T.A. Nelson, Y.C. MacNab, A.B. Lawson, Review of methods for space-time disease surveillance, *Spatial and Spatio-temporal Epidemiology* 1 (2010) 105–116.
- [43] V. Sintchenko (Ed.), *Infectious Disease Informatics*, Springer, New York, 2010.
- [44] C. Sonesson, D. Bock, A review and discussion of prospective statistical surveillance in public health., *Journal of the Royal Statistical Society, Series A* 166 (1) (2003) 5–21.
- [45] Statistics Canada. <http://www12.statcan.gc.ca/census-recensement/2006/>
- [46] C.L. Vidal Rodeiro, A.B. Lawson, Monitoring changes in spatio-temporal maps of disease, *Biometrical Journal* 48 (3) (2006) 463–480.
- [47] M.M. Wagner, A.W. Moore, R.M. Aryel, *Handbook of Biosurveillance*, Elsevier, London, 2006.
- [48] R. Watkins, S. Eagleson, B. Veenendaal, G. Wright, A.J. Plant, Disease surveillance using a hidden Markov model, *BMC Medical Informatics and Decision Making* 9 (1) (2009) 39.